Optimal Radio Access Technology Selection Algorithm for LTE-WiFi Network [†]

Arghyadip Roy, *Student Member, IEEE,* Prasanna Chaporkar, *Member, IEEE,*
and Abhay Karandikar, *Member, IEEE.*

Abstract—**Heterogeneous Network (HetNet) comprises of multiple Radio Access Technologies (RATs) allowing a user to associate with a specific RAT and steer to other RATs in a seamless manner. To cope up with the unprecedented growth of data traffic, mobile data can be offloaded to Wireless Fidelity (WiFi) in a Long Term Evolution (LTE) based HetNet. In this paper, an optimal RAT selection problem is considered to maximize the total system throughput in an LTE-WiFi system with offload capability. Another formulation is also developed where maximizing the total system throughput is subject to a constraint on the voice user blocking probability. It is proved that the optimal policies for the association and offloading of voice/data users contain threshold structures. As a policy search over the entire policy space may become computationally inefficient, we propose computationally efficient algorithms based on the threshold structures for the association and offloading of users in LTE-WiFi HetNet. Simulation results are presented to demonstrate the voice user blocking probability and the total system throughput performance of the proposed algorithms in comparison to other benchmark algorithms.**

Index Terms—**User association, LTE-WiFi offloading, CMDP, Threshold policy.**

I. INTRODUCTION

To meet the ever-increasing Quality of Service (QoS) requirements of users, various Radio Access Technologies (RATs) have been standardized [1]-[2]. Each RAT has different characteristics regarding associated parameters like coverage and capacity. It has been predicted that by 2021 monthly global mobile data traffic will exceed 49 exabytes [3]. This unprecedented growth in data traffic has become one of the serious challenges for cellular network operators. To address this issue, both from users' and network providers' point of view, it has become necessary that different RATs interwork with each other. A wireless network where different RATs are present, and users can be associated and moved seamlessly from one RAT to another, is called a Heterogeneous Network (HetNet). In this paper, our aim is to determine the optimal RAT selection policy in a HetNet [1].

Due to the complementary characteristics of Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) Base Stations (BSs) providing ubiquitous coverage and IEEE 802.11 [4] based Wireless Local Area Network (WLAN) (also known as Wireless Fidelity (WiFi)) Access Points (APs) providing high bit rate capability in hot-spot areas, interworking between them [5] offers an interesting solution. In areas where both LTE and WiFi coverage are present, a user can be associated with either of them. Moreover, data users can be steered from one RAT to another to achieve load balancing. This proposal, known as mobile data offloading, has been introduced in 3GPP Release 12 specifications [5]. Since WiFi operates in unlicensed spectrum and most of the commercially available user equipment already have a dedicated WLAN interface, this proposal has become popular both with network operators and handset manufacturers.

For efficient utilization of both LTE and WiFi networks, it is necessary to take appropriate association and offloading decisions. RAT selection and offloading decisions can be made either at the user side or the network side. In user-initiated RAT selection schemes, there is no cooperation between LTE and WiFi networks, and users decide which RAT should be selected based on certain criteria. Since users individually take selfish RAT selection decisions to maximize individual utility functions, this may not provide a globally optimum solution [6]-[10]. To address this issue, a network-initiated RAT selection algorithm, which optimizes different network parameters, becomes necessary.
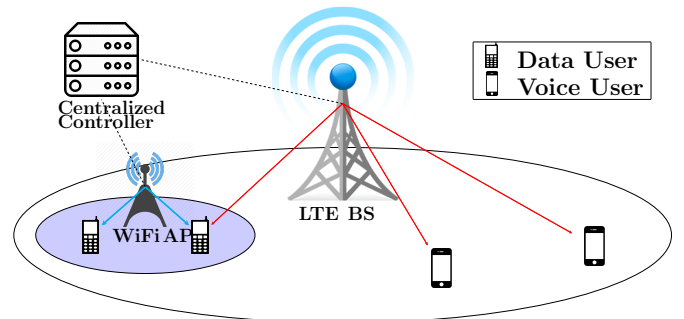

Figure 1: LTE-WiFi heterogeneous network architecture.

In this paper, we investigate an optimal association policy for an LTE-WiFi HetNet, as illustrated in Fig.1. Network-initiated RAT selection and offloading decisions are taken by a centralized controller possessing an overall view of the network. We consider two types of users, viz., voice and data users, to be present inside the LTE-WiFi HetNet. We consider that voice users are always associated with LTE since unlike LTE, WiFi may not provide the required QoS for a voice user. However, data users can be associated with either LTE or WiFi. Offloading of data users from one RAT to another is considered at the time of association of voice

[1]The terminologies "RAT selection" and "association" have been used interchangeably throughout the paper.

users or departure of existing voice/data users. From a network operator's perspective, total system throughput is an important system metric since the generated revenue may largely depend on the number of bytes transported by the operator. Moreover, data users experiencing high throughput are more likely to adhere to a network operator, thus facilitating the improvement of the customer base of the operator. Therefore, we aim to maximize the total system throughput and formulate this as a continuous-time Markov Decision Process (MDP) problem.

In the case of data users, although in low load condition, WiFi usually provides higher throughput than that of LTE, as the WiFi load increases, average per-user throughput in WiFi decreases rapidly [11]. Therefore, under high WiFi load, LTE may offer more throughput than WiFi to data users and thus may be preferable to data users for the association. However, voice and data users are allocated resources in LTE from a common resource block pool. The throughput requirement of LTE data users is usually more than that of the voice users. Therefore, maximization of the total system throughput may result in excessive blocking of voice users. The system may attempt to save LTE resources which can be allocated later to data users having greater contributions to the system throughput than that of voice users. It results in an inherent trade-off between the total system throughput and the blocking probability of voice users. We consider this problem within the formalism of Constrained Markov Decision Process (CMDP), which maximizes the total system throughput subject to a constraint on the voice user blocking probability.

It is proved that the associated optimal policies contain a threshold structure, where after a certain threshold on the number of WiFi data users, data users are served using LTE. The existence of a similar threshold for the blocking of voice users is also established. Based on the threshold based optimal policy, we propose two RAT selection algorithms for LTE-WiFi HetNet. We derive that by virtue of the threshold properties, a significant reduction in computational complexity is achieved over the traditional policy iteration [12] algorithm to obtain the optimal policy. Extensive simulations are performed in ns-3 (a discrete event network simulator) [13] to evaluate the performance of the proposed association algorithms. Using simulation results, performance gains of the proposed algorithms in comparison to other algorithms in the literature e.g.,[14] are also evaluated.

A. Related Work

The solutions which investigate RAT selection problem in a HetNet can be broadly divided into two categories, viz., user-initiated [6]-[10] and network-initiated [15]-[27]. In [6], a user-initiated RAT selection algorithm based on Signal-to-Noise Ratio (SNR) and load information of individual RATs with the adaptation of hysteresis mechanism, is considered for LTE-WiFi HetNet. The performance of this scheme is compared with network-initiated cell-range extension schemes that use network-optimized Received Signal Strength Indicator (RSSI) bias value to steer users to other RATs. In [7], a distributed RAT selection algorithm is proposed based on the distance and peak rate obtained from different IEEE 802.11 [4] APs. Authors in [8] propose an association algorithm in an LTE-WiFi HetNet to address the trade-off between QoS and resource utilization.

Few heuristic-based network-initiated RAT selection approaches are considered in [23]-[25]. While the algorithm proposed in [24] prefers WLAN over cellular regardless of the service type, the one proposed in [25] prefers cellular RAT for voice users and WLAN for data users. Among the other network-initiated RAT selection schemes, [15]-[22], [26]-[27] consider various optimization approaches. In [15], optimal RAT selection problem is addressed in a HetNet to optimize throughput, blocking probability, etc.. Since the associated algorithm scales exponentially with the system size, authors also propose a computationally efficient heuristic policy. In [16], the association resulting in maximum value for the sum of logarithms of throughputs is chosen as the optimal association among Wireless Stations (STAs) and APs. However, authors do not take into account user arrival and departure. A context-aware RAT selection algorithm for a HetNet is proposed in [17], which operates on the user side with assistance from the network and aims to minimize the signaling overhead and the amount of computation at the base stations. A joint user association and interference management problem in a two-tier HetNet is considered in [18]. Authors introduce a pico operation mode and propose a computationally efficient method to maximize the network utility under proportional fairness. However, this approach is not adaptable to quick changes in network dynamics. RAT selection policies in wireless networks [28]-[30] are sometimes observed to contain certain threshold structures. A multi-class admission system is considered in [28], where it is demonstrated that if it is optimal to accept a user of a class, then it is optimal to accept a user of higher profit class too.

Offloading of data users [1]-[2] from one RAT to another plays a major role in the capacity improvement of the system. Performance improvement achieved by on-the-spot offloading [31], a user-initiated WiFi offloading scheme, is analyzed in [14]. The basic idea behind on-the-spot offloading is to steer the mobile data users to WiFi, whenever WiFi is available. Authors in [32] consider the trade-off between the cost and delay involved in opportunistic WiFi offloading and propose an adaptive strategy to maximize the user's satisfaction. The user-initiated offloading scheme in [33] is based on the combined information of signal strength and network load of LTE/WLAN. However, being a greedy one, this algorithm fails to converge to a globally optimum solution. The network-initiated offloading approach in [34] computes the optimal fraction of traffic to be offloaded to WiFi such that the per-user throughput of the system is maximized and performs better than on-the-spot offloading [14]. However, the model in [34] does not incorporate voice users inside an LTE network.

B. Our Contribution

In this paper, we investigate the optimal association policy in an LTE-WiFi HetNet. We consider a system where voice and data users can arrive or depart at any point in time. We introduce the possibility of data user offloading from one RAT to another at the time of association or departure of a user. Our contribution can be summarized as follows.

- We target to maximize the total system throughput. The problem is formulated within the framework of MDP.

To address the inherent trade-off between the total system throughput and the blocking probability of voice users, another formulation is developed, where we target to maximize the total system throughput, subject to a constraint on the voice user blocking probability, using CMDP.

- Threshold structures of optimal policies are established.
- We propose two algorithms based on the threshold structures of optimal policies along with an analysis of computational complexity and implement in ns-3. 3GPP recommended parameters are used in the simulations.
- Since most of the practical offloading schemes offload data users to WiFi, performances of the proposed algorithms are compared with on-the-spot offloading algorithm [14]. Furthermore, we compare the performance of our schemes with the "LTE-preferred" scheme, where data users prefer LTE over WiFi.

The arrival of a new user in the LTE-WiFi system triggers the need for the optimal RAT selection. Also, with the arrival or departure of users, the active users in different RATs may need to get offloaded to other RATs. While few works in the literature have focused on RAT selection and offloading techniques, respectively, no existing literature, to the best of our knowledge, has addressed the issue of joint RAT selection and offloading for LTE-WiFi HetNet. Traditional algorithms for obtaining the optimal policy search over the entire policy space and hence become computationally inefficient. The algorithms proposed by us exploit the threshold nature of optimal policy derived by us and offers significantly lower computational complexity to determine the optimal policy.

The rest of the paper is organized as follows. The system model is described in Section II. In Section III, the RAT selection problems are formulated within the framework of unconstrained and constrained continuous-time MDP, respectively. In Section IV, we derive the threshold structure of the optimal policy. Algorithms for the association of voice and data users in LTE-WiFi HetNet are proposed in Section V. Section VI presents simulation results. In Section VII, we conclude the paper.

II. System Model

We consider a system where an LTE BS and a WiFi AP are present. As illustrated in Fig.1, we assume that both the BS and the AP are connected to a centralized controller by lossless links. We assume that the voice and data users are geographically located at any point in the LTE BS coverage area. Since data users outside the dual coverage area of the LTE BS and the WiFi AP always get associated with the LTE BS and no decision is involved in this case, without loss of generality, we take into consideration only those data users which are present inside the WiFi AP coverage area. We assume that there is a common resource pool in LTE for the voice users as well as the data users inside the WiFi AP coverage area. Data users inside the dual coverage area can be associated with the LTE BS or the WiFi AP. All the users are assumed to be stationary. Voice and data user arrivals follow Poisson processes with means λ_v and λ_d, respectively. Service times for voice and data user are exponentially distributed with

means $\frac{1}{\mu_v}$ and $\frac{1}{\mu_d}$, respectively. For justification behind these assumptions, see [35].

Remark 1. *Although for the brevity of notation, a single LTE BS and a single AP have been considered, the system model can be generalized to a single LTE BS and multiple APs with non-overlapping coverage areas by considering the number of data users in different APs in the system model. Moreover, considering that each point in a geographical area is mapped to a single LTE BS (the LTE BS with highest average signal strength, say), multiple BSs can also be included in the system model with slight modifications. In the case of multiple overlapping APs inside the coverage area of an LTE BS, a one-to-one mapping between a user location and an AP using a decision criterion (based on highest average signal strength, say) reduces the problem to a single BS-multiple non-overlapping AP problem. The set where more than one AP have the same average signal strength is a thin set (Lebesgue measure=0).*

Remark 2. *Most of the cellular network operators support seamless offloading of data users to trusted (operator-deployed) WiFi APs. The decision regarding when to offload is at the discretion of the service providers. We build our system model based on the architecture provided by cellular operators for interworking with WiFi and propose optimal algorithms for RAT selection and offloading.*

A. State Space

We model the system as a controlled continuous time stochastic process $\{X(t)\}_{t \geq 0}$ defined on a state space $\mathcal{S}$. Any state $s \in \mathcal{S}$ is represented as a 3-tuple $s = (i, j, k)$, where i, j and k represent the number of voice users in LTE, the number of data users in LTE and the number of data users in WiFi, respectively. The system state remains unchanged unless an existing user departs or a new user arrives in the system. The arrivals and departures in the system are referred to as events. Five types of events are possible, viz., (E_1) an arrival of a new voice user in the system, (E_2) an arrival of a new data user in the system, (E_3) a departure of an existing voice user from LTE, (E_4) a departure of an existing data user from LTE and (E_5) a departure of an existing data user from WiFi. Whenever an event occurs, the centralized controller takes an action, and based on the type of event and the action taken by the controller, a state transition may happen. Note that the transitions of $\{X(t)\}_{t \geq 0}$ happen only at the event epochs and not otherwise. Thus, it suffices to observe the system state only at event epochs. A finite amount of reward and cost are associated with every state-action pair. Detailed descriptions of the action space, state transitions, reward, and cost are provided in subsequent subsections.

Next, we elaborate on the structure of $\mathcal{S}$. We assume that $(i, j, k) \in \mathcal{S}$ if $(i + j) \leq C$ and $k \leq W$, where C is the total number of common resource blocks reserved in LTE for voice and data users, and W is the maximum number of users in WiFi, so that the per-user throughput in WiFi is greater than a threshold. The condition $(i + j) \leq C$ arises because we assume that in each LTE subframe, every admitted user is allocated one resource block. If this allocation is not possible, a new user is not admitted in the LTE system. Furthermore,

note that WiFi throughput decays monotonically [11] as the number of WiFi users increases. We assume that each user gets more than a threshold value of average throughput (say 2 Mbps), which leads to the bound W on the maximum number of users that can be accommodated in the WiFi system.

Remark 3. *Although the allocation of multiple resource blocks is closer to the practical scenario, this complicates the system model while the methodology and approach adopted in this paper do not change.*

B. Action Space

The set of actions defines a set of possible association and offloading strategies in the event of arrival or departure of a user. Let the action space be denoted by $\mathcal{A}$. Depending on the arrival or departure, we have a set of actions as stated below.

$$\mathcal{A} = \begin{cases} A_1, & \text{Block the arriving user or do nothing} \\ & \text{during departure,} \\ A_2, & \text{Accept voice/data user in LTE,} \\ A_3, & \text{Accept data user in WiFi,} \\ A_4, & \text{Accept voice user in LTE and offload} \\ & \text{one data user to WiFi,} \\ A_5, & \text{Move one data user to a RAT (from} \\ & \text{which departure has occurred).} \end{cases}$$

Remark 4. *In this paper, actions are chosen based on the system state and the event occurred. One way of representing this is embedding the event in the state space so that the action depends only on the system state. However, to avoid notational complications associated with this approach, we view the action as a function of the system state and the event.*

Let the set of states (subset of $\mathcal{S}$) in which action a chosen based on an event E_l is feasible be denoted by $\mathcal{S}_{E_l,a}$. The complete description of $\mathcal{S}_{E_l,a}$ for different events and actions is provided in [36]. In the case of voice and data user arrivals, the set of all possible actions are $\{A_1, A_2, A_4\}$ and $\{A_1, A_2, A_3\}$, respectively. However, when an event E_l occurs, action a is not feasible if the system state is not present in $\mathcal{S}_{E_l,a}$. In this paper, voice user blocking (A_1) is considered to be a feasible action in all the states, provided the system is not empty. We consider blocking as a feasible action for data users, only when capacity is reached for both the RATs. When a user departs from LTE or WiFi, the controller can choose either A_1 or A_5. If after the departure of a user from LTE, A_5 is chosen, it offloads one data user from WiFi to LTE.

C. State Transitions

Based on an event and an action chosen, from a state, the system moves deterministically to a different state. Assume that from the state $s = (i, j, k)$, the system moves to the state $s'(E_l, a) = (i', j', k')$ under the event E_l and chosen action a. Values of i', j' and k' for different events E_l (arrivals and departures of users) and action a are tabulated in Table I. Note that this table is exhaustive in all kinds of events and actions. However, in a state, we need to consider only those events and actions which are feasible in that state. Fig.2 describes state transitions from the state (i, j, k) under different actions corresponding to various arrival and departure events, respectively. In the diagram, the transition under action a corresponding to event E_l is denoted by $a|E_l$.

Table I: Transition Probability Table.

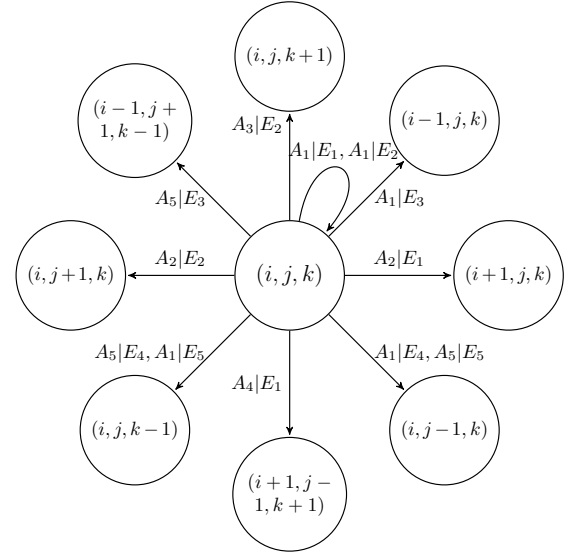| (E_l, a)	(i', j', k')
(Voice arrival(E_1), A_1)	(i, j, k)
(Data arrival(E_2), A_1)	(i, j, k)
(Voice departure from LTE(E_3), A_1)	$(i-1, j, k)$
(Data departure from LTE(E_4), A_1)	$(i, j-1, k)$
(Data departure from WiFi(E_5), A_1)	$(i, j, k-1)$
(Voice arrival(E_1), A_2)	$(i+1, j, k)$
(Data arrival(E_2), A_2)	$(i, j+1, k)$
(Data arrival(E_2), A_3)	$(i, j, k+1)$
(Voice arrival(E_1), A_4)	$(i+1, j-1, k+1)$
(Voice departure from LTE(E_3), A_5)	$(i-1, j+1, k-1)$
(Data departure from LTE(E_4), A_5)	$(i, j, k-1)$
(Data departure from WiFi(E_5), A_5)	$(i, j-1, k)$

Figure 2: State transition diagrams under different actions and events.

D. Rewards and Costs

Let the reward and cost functions per unit time corresponding to a state s, event E_l and action a be represented by $r(s, E_l, a)$ and $c(s, E_l, a)$, respectively. Let $R_{L,V}$ and $R_{L,D}$ denote the bit rate of voice and data users in LTE, respectively. To keep the model simple and computationally tractable, we assume that the bit rate of data users in LTE is constant. In general, a voice user generates constant bit rate (CBR) traffic, and hence we take $R_{L,V}$ to be a constant. $R_{W,D}(k)$ corresponds to the per-user data throughput of k users in WiFi. We assume full buffer traffic model [11] for WiFi. The calculation of $R_{W,D}(k)$ is based on the contention-driven medium access of WiFi users. It is a function of the probabilistic transmission attempts of the users, corresponding success and collision probabilities, and slot times for successful transmission, idle slots and busy slots during collisions.

Remark 5. *Data users want to experience the maximum possible data rate. We assume that a Transmission Control Protocol (TCP) connection is established between the data user and the LTE BS (WiFi AP) before the data transfer begins. The throughput term signifies the maximum data rate provided by the TCP pipe for a single TCP connection.*

The reward per unit time in a state under the occurrence of an event and an action chosen is defined as the total system throughput in that state under that event and the chosen action. The complete description of reward rates for different events and actions in state s is provided in Table II. Note that the

reward rate is a monotone increasing function of i and j.

Table II: Reward Rate Table.

(E_l, a)	$r(s, E_l, a)$
(E_1, A_1)	$iR_{L,V} + jR_{L,D} + kR_{W,D}(k)$
(E_2, A_1)	$iR_{L,V} + jR_{L,D} + kR_{W,D}(k)$
(E_3, A_1)	$(i-1)R_{L,V} + jR_{L,D} + kR_{W,D}(k)$
(E_4, A_1)	$iR_{L,V} + (j-1)R_{L,D} + kR_{W,D}(k)$
(E_5, A_1)	$iR_{L,V} + jR_{L,D} + (k-1)R_{W,D}(k-1)$
(E_1, A_2)	$(i+1)R_{L,V} + jR_{L,D} + kR_{W,D}(k)$
(E_2, A_2)	$iR_{L,V} + (j+1)R_{L,D} + kR_{W,D}(k)$
(E_2, A_3)	$iR_{L,V} + jR_{L,D} + (k+1)R_{W,D}(k+1)$
(E_1, A_4)	$(i+1)R_{L,V} + (j-1)R_{L,D} + (k+1)R_{W,D}(k+1)$
(E_3, A_5)	$(i-1)R_{L,V} + (j+1)R_{L,D} + (k-1)R_{W,D}(k-1)$
(E_4, A_5)	$iR_{L,V} + jR_{L,D} + (k-1)R_{W,D}(k)$
(E_5, A_5)	$iR_{L,V} + (j-1)R_{L,D} + kR_{W,D}(k)$

The cost function considered here is as follows. Whenever the centralized controller blocks an incoming voice user, one unit cost is incurred per unit time. Otherwise it is zero. Thus,

$$c(s, E_l, a) = \begin{cases} 1, & \text{if voice user is blocked,} \\ 0, & \text{else.} \end{cases}$$

We consider the blocking of data users only when both the LTE and the WiFi systems are full. Hence, we do not consider any cost on the blocking of data users. Note that once a voice user is associated with LTE, QoS in terms of delay and data rate is guaranteed by allocating dedicated bearers providing Guaranteed Bit Rate (GBR) in LTE. Therefore, in this paper, we consider only the blocking probability as the QoS requiremet for voice users.

III. PROBLEM FORMULATION AND SOLUTION METHODOLOGY

A *decision rule* describes the mapping regarding which action is to be chosen at different states $s \in S$ and decision epochs t_n. An association *policy* is a sequence of decision rules $(\pi^{t_1}, \pi^{t_2}, \ldots, \pi^{t_n}, \ldots)$ taken at different decision epochs. Our goal is to determine an association policy which maximizes the total system throughput. This can be formulated as a continuous-time unconstrained MDP problem. In this case, a *pure optimal policy* exists [12]. Since the contribution of data users to the total system throughput is more than that of the voice users, the optimal association policy may result in high blocking probability of voice users. Hence, to address the trade-off between the total system throughput and the voice user blocking probability, we consider the CMDP problem, where we target to maximize the total system throughput, subject to a constraint on the voice user blocking probability. In this case, a stationary randomized optimal policy exists [37]. A *Randomized policy* is a mixture of two pure policies with associated probabilities. Arrival and departure of users can occur at arbitrary points in time, which makes the problem continuous time in nature.

A. Problem Formulation

Let $\mathcal{M}$ be the set of all memoryless policies. To guarantee a unique stationary distribution, we assume that Markov chains associated with such policies are irreducible. Following the policy $M \in \mathcal{M}$, let the average reward and cost of the system over infinite horizon be denoted by V^M and B^M, respectively. Let $R(t)$ and $C(t)$ be the total reward and cost of the system incurred up to time t, respectively. For the unconstrained

MDP problem, our objective is to maximize the total system throughput which can be described as follows.

$$\text{Maximize:} \quad V^M = \lim_{t \to \infty} \frac{1}{t} \mathbb{E}_M[R(t)], \tag{1}$$

where $\mathbb{E}_M$ denotes the expectation operator under the policy M. However, for the CMDP problem, our objective is to maximize the total system throughput, subject to a constraint on the blocking probability of voice users. This can be described as follows.

$$\begin{aligned} \text{Maximize:} \quad & V^M = \lim_{t \to \infty} \frac{1}{t} \mathbb{E}_M[R(t)], \\ \text{Subject to:} \quad & B^M = \lim_{t \to \infty} \frac{1}{t} \mathbb{E}_M[C(t)] \leq B_{\max}, \end{aligned} \tag{2}$$

where $B_{\max}$ denotes the constraint on the blocking probability of voice users. Our objective is to determine the optimal policy for both unconstrained and constrained MDP problem. Since the optimal policies are known to be stationary policies, the corresponding limits in Equation (1) and (2) exist.

B. Conversion to Discrete-Time MDP

To compute the optimal policy, we can use the well-known Value Iteration Algorithm (VIA) [12]. However, before that, the continuous-time MDP has to be transformed into an equivalent discrete-time MDP using uniformization [12], so that both models have the same average expected reward and cost for a stationary policy.

Let $\tau_s(E_l, a)$ represent the expected time until the next event, if action a is chosen in state s under the event E_l. We need to choose a number δ, such that $0 < \delta \ll \min_{s, E_l, a} \tau_s(E_l, a)$. The state space and the action space remain the same in the equivalent discrete-time model. Let $\hat{p}(s, E_l)$, $\hat{r}(s, E_l, a)$ and $\hat{c}(s, E_l, a)$ represent the probabilities of the event, reward and cost in the transformed discrete-time model in state s under the event E_l and action a, respectively. Thus, we have,

$$\hat{r}(s, E_l, a) = r(s, E_l, a) \quad \text{and} \quad \hat{c}(s, E_l, a) = c(s, E_l, a).$$

We have, $\hat{p}(s, E_1) = \lambda_v \delta$, $\hat{p}(s, E_2) = \lambda_d \delta$, $\hat{p}(s, E_3) = i\mu_v\delta$, $\hat{p}(s, E_4) = j\mu_d\delta$ and $\hat{p}(s, E_5) = k\mu_d\delta$. Note that, this discrete-time MDP has identical stationary policies to that of the continuous-time MDP.

C. Lagrangian Approach

After conversion into an equivalent discrete-time MDP model, we use the Lagrangian approach [37] to solve the CMDP. For a fixed value of Lagrange Multiplier (LM) β, the modified reward function of the CMDP is

$$\hat{r}(s, E_l, a; \beta) = \hat{r}(s, E_l, a) - \beta\hat{c}(s, E_l, a).$$

The dynamic programming equation below describes the necessary condition for optimality.

$$\begin{aligned} V(s) = & \sum_{E_l} \hat{p}(s, E_l) \max_a [\hat{r}(s, E_l, a; \beta) + V(s'(E_l, a))] \\ & + \left(1 - \sum_{E_l} \hat{p}(s, E_l)\right) V(s), \end{aligned} \tag{3}$$

where $V(s)$ denotes the value function in state $s \in \mathcal{S}$. Next, our aim is to determine the value of $\beta (= \beta^*$, say) which maximizes the average expected reward, subject to a cost

constraint. The value of β^* can be determined using gradient descent algorithm, as discussed in [38]. In kth iteration, we modify the value of β from its previous iteration as,

$$\beta_{k+1} = \beta_k + \frac{1}{k}(B^{\pi_{\beta_k}} - B_{\max}). \tag{4}$$

For a fixed value of β, the unconstrained maximization problem can be solved using VIA, as described below.

$$V_{n+1}(s) = \sum_{E_l} \hat{p}(s, E_l) \max_a [\hat{r}(s, E_l, a; \beta) + V_n(s'(E_l, a))]$$
$$+ \Big(1 - \sum_{E_l} \hat{p}(s, E_l)\Big) V_n(s), \tag{5}$$

where $V_n(.)$ is an estimate of the value function after nth iteration. After determining β^*, the next step is to determine the optimal policy for the CMDP problem. As discussed in [37], optimal policy for a CMDP problem is a mixture of two pure policies $\pi_{\beta^*-\epsilon}$ and $\pi_{\beta^*+\epsilon}$, obtained by perturbation of β^* by a small amount ϵ in both directions. Let the long-term average expected costs of the two pure policies be $B_{\beta^*-\epsilon}$ and $B_{\beta^*+\epsilon}$, respectively. In the next step, we determine the value of the parameter p such that

$$pB_{\beta^*-\epsilon} + (1-p)B_{\beta^*+\epsilon} = B_{\max}.$$

Finally, we have a randomized optimal policy for the considered CMDP problem. At each decision epoch, policies $\pi_{\beta^*-\epsilon}$ and $\pi_{\beta^*+\epsilon}$ are chosen w.p. p and $(1-p)$, respectively.

Note that the iterations on LM described above are necessary only for the CMDP problem. In the case of unconstrained MDP, VIA can be employed to determine a pure optimal policy, after an equivalent discrete-time MDP model is obtained.

IV. STRUCTURE OF THE OPTIMAL POLICY

The dynamic programming equations (Equation (3) and (5)) described in the previous section are exploited to establish the fact that the optimal policy is of threshold type. The optimality of threshold policy is established with the aid of some lemmas presented below. For the purpose of readability, we present the proofs of the lemmas in Appendices.

A. Optimal Policy for Data Users

In this section, we present structural properties on the optimal policy for the service of data users along with their physical interpretations. Let us denote the throughput increment in WiFi when the number of WiFi users increases from k to $(k+1)$ by $\tilde{R}_{W,D}(k)$. Therefore, $\tilde{R}_{W,D}(k) = (k+1)R_{W,D}(k+1) - kR_{W,D}(k)$. We assume the following.

Assumption 1. *Let $R_{L,D}$ be such that $R_{L,D} \geq \tilde{R}_{W,D}(k)$, $\forall k \geq k_{th}$ and $R_{L,D} < \tilde{R}_{W,D}(k), \forall k < k_{th}$, where k_{th} is a threshold such that if $k \geq k_{th}$, the data rate improvement provided by a single data user in the LTE system is more than the improvement in total WiFi throughput as the number of WiFi data users is increased from k to $(k+1)$.*

Remark 6. *Following the full buffer traffic model [11], $\tilde{R}_{W,D}(k)$ initially increases with k and then decreases. This behavior matches with Assumption 1. The value of k_{th} can be easily obtained by computing $\tilde{R}_{W,D}(k)$ (following [11]) for different values of k and comparing with $R_{L,D}$.*

The following two lemmas describe a threshold structure on the optimal policy for the service of data users. Specifically, up to a certain threshold on the total number of data users, data users are served using WiFi. After the threshold is crossed, data users are served using LTE.

Lemma 1. *For every i and j such that $(i+j) < C$, if the total number of data users in the system is $(j+k) \leq k_{th}$, then the optimal policy is to serve all data users using WiFi. In other words, $(j+k) \leq k_{th} \implies j = 0$.*

Proof. See Apppedix A. $\square$

In Lemma 1, following Assumption 1, since for $k < k_{th}$, the data rate improvement is more if an additional data user is served using WiFi rather than using LTE, it is optimal to serve the data users using WiFi.

Lemma 2. *For every i and j such that $(i+j) < C$, if the total number of data users in the system is $(j+k) > k_{th}$, then the optimal policy is to serve k_{th} data users using WiFi and all other data users using LTE. In other words, $(j+k) > k_{th} \implies k = k_{th}$.*

Proof. See Apppedix B. $\square$

The physical significance of Lemma 2 is that for $k \geq k_{th}$, the data rate improvement provided by a single data user in LTE is more than that of the WiFi (following Assumption 1), and hence it is optimal to serve up to k_{th} data users using WiFi and serve the additional data users using LTE.

Following lemma is a direct consequence of how the system is modeled.

Lemma 3. *For every i and j such that $(i+j) = C$, the optimal policy is to serve all the incoming data users using WiFi until $k = W$, where an incoming data user is blocked.*

B. Optimal Policy for Voice Users

In this section, we characterize the optimal policy for the arrival of voice users. We prove that the optimal policy is of threshold type. Let D_i be the difference operator which is defined as $D_i V(i,j,k) = V(i+1,j,k) - V(i,j,k)$. Similarly, we define the second difference operator as $D_{ii}(.) = D_i(D_i(.))$. Let E_i be another difference operator defined as $E_i V(i,j,k) = V(i+1, j-1, k+1) - V(i,j,k)$. We define the second difference operator as $E_{ii}(.) = E_i(E_i(.))$. Similarly, we define $F_i V(i,j,k) = V(i+1, j-1, k) - V(i,j,k)$. In this section, the terminologies "increasing" and "decreasing" are used in the weak sense of "non-decreasing" and "non-increasing", respectively. In each state, let the sum of arrival and service rates be denoted by $v(i,j,k)$. Thus, we have,

$$v(i,j,k) = \lambda_v + \lambda_d + i\mu_v + j\mu_d + k\mu_d.$$

Let us define $f(i,j,k) = (iR_{L,V} + jR_{L,D} + kR_{W,D}(k))$. The lemma presented below describes the superiority of one action over the other for the service of incoming voice users. Specifically, up to a certain threshold on the total number of data users, A_4 (accept voice user in LTE with data user offload to WiFi) is better than A_2 (accept voice user in LTE). After the threshold is crossed, A_2 becomes better.

Lemma 4. *In the case of a voice user arrival in a state (i, j, k), where $(i + j) < C$,*

(i) *A_4 is always better than A_2 if $k < k_{th}$,*
(ii) *A_2 is always better than A_4 if $k \geq k_{th}$.*

Proof. Proof is similar to the proof of Lemma 1. $\qquad\square$

Similar to Lemmas 1 and 2, following Assumption 1, since for $k < k_{th}$, the data rate improvement is more if an additional data user is served in WiFi rather than in LTE, A_4 is better than A_2. Hence, when $k < k_{th}$, the choice of optimal action is between A_4 (accept voice user in LTE with data user offload to WiFi) and A_1 (blocking). Similarly, for $k \geq k_{th}$, the optimal action is either A_2 (accept voice user in LTE) or A_1.

The following lemma describes that when capacity is not reached in LTE and a voice user arrives, a threshold structure is observed. Until a threshold on the number of voice/data users in LTE, A_2 (for $k \geq k_{th}$) or A_4 (for $k < k_{th}$) is preferred. After the threshold A_1 becomes optimal.

Lemma 5. *For every i and j such that $(i + j) < C$ and a voice user arrival,*

(i) *if the optimal action in state (i, j, k) is A_1, then the optimal action in state $(i+1, j, k)$ and in state $(i, j+1, k)$ is also A_1,*
(ii) *if the optimal action in state (i, j, k) is A_2 (A_4), then the optimal action in state $(i-1, j, k)$ and in state $(i, j-1, k)$ is also A_2 (A_4).*

Proof. See Apppedix C. $\qquad\square$

When the number of voice/data users in LTE is less, A_2 or A_4 is chosen as the optimal action in the event of a voice user arrival. When i or j crosses a certain threshold, the number of free resources for incoming voice users decreases. Therefore, the blocking probability of voice users increases. Thus, after a threshold on i or j, A_1 becomes optimal.

However, when $(i + j) = C$, since A_2 is infeasible, optimal action is either A_1 or A_4. The lemma presented below discusses the threshold nature of the optimal policy for voice user arrivals when $(i + j) = C$.

Lemma 6. *For every i and j such that $(i + j) = C$ and a voice user arrival,*

(i) *if the optimal action in state (i, j, k) is A_1, then the optimal action in state $(i + 1, j - 1, k)$ is also A_1,*
(ii) *if the optimal action in state (i, j, k) is A_4, then the optimal action in state $(i - 1, j + 1, k)$ is also A_4.*

Proof. See Apppedix D. $\qquad\square$

The physical interpretation of the above lemma is that for states with $(i + j) = C$, when i is small, A_4 is preferred. However, when i crosses a threshold, since j becomes small, and consequently the total system throughput is small, A_4 may further lower the total system throughput. Therefore, blocking of voice users is chosen as the optimal action.

V. Proposed Network-Initiated Association Algorithms

Based on the threshold structures of the optimal policies for the unconstrained MDP and CMDP problems, in this section,

---

Algorithm 1 Network-Initiated Unconstrained MDP-based Association Algorithm.

Input: $\lambda_v, \lambda_d, \mu_v, \mu_d, R_{L,V}, R_{L,D}, R_{W,D}(.)$.
 1: Compute k_{th} for data users (See Remark 6).Set $\beta \leftarrow 0$.
 2: **procedure** Threshold−Policy−Search
 3:     Initialize threshold $th_0 \leftarrow th$. Set $n \leftarrow 0$.
 4:     **procedure** Policy−Evaluation
 5:         Solve the linear system for policies corresponding to thresholds $th_n, (th_n - 1)$ and $(th_n + 1)$.
 6:         $V(s) = \sum_{E_l} \hat{p}(s, E_l)[\hat{r}(s, E_l, M(s); \beta) + V(s'(E_l, M(s)))] + (1 - \sum_{E_l} \hat{p}(s, E_l))V(s), \forall s \in \mathcal{S}$.
 7:     **end procedure**
 8:     **procedure** Policy−Improvement
 9:         If $V_{th_n+1} \succeq V_{th_n}$, set $th_{n+1} \leftarrow th_n + 1$.
10:         Elseif $V_{th_n-1} \succeq V_{th_n}$, set $th_{n+1} \leftarrow th_n - 1$.
11:         Else set $th_{n+1} \leftarrow th_n$.
12:         If $th_{n+1} == th_n$ stop,
13:         Else $n \leftarrow n + 1$ and go to Line 4.
14:     **end procedure**
15: **end procedure**
Output: Deterministic optimal policy.
16: Store thresholds $va_c(j, k)$ and $va_{lc}(j, k)$ for the association of voice users for $(i + j) = C$ and $(i + j) < C$, respectively.
17: **procedure** Policy−Impl
18:     **for** each arrival of voice users **do**
19:         **if** $(i + j) < C$ **then**
20:             Choose A_1 if $i \geq va_{lc}(j, k)$.
21:             Choose A_2 if $i < va_{lc}(j, k)$ and $k \geq k_{th}$.
22:             Choose A_4 otherwise.
23:         **else**
24:             Choose A_4 if $i < va_c(j, k)$, A_1 otherwise.
25:         **end if**
26:     **end for**
27:     **for** each arrival of data users **do**
28:         **if** $(i + j) < C$ **then**
29:             Choose A_3 if $k < k_{th}$, A_2 otherwise.
30:         **else**
31:             Choose A_3 if $k < W$, A_1 otherwise.
32:         **end if**
33:     **end for**
34:     **for** each departure of users from LTE (WiFi) **do**
35:         Choose A_1 (A_5) if $k \leq k_{th}$, $A_5(A_1)$ otherwise.
36:     **end for**
37: **end procedure**

---

we propose two network-initiated association algorithms for LTE-WiFi HetNet.

A. Unconstrained MDP-based Association Algorithm

The details of the unconstrained MDP-based algorithm is presented in Algorithm 1. As discussed in Section IV, the threshold for the service of data users k_{th} can be computed easily without the knowledge of arrival and service rates. The value of k_{th} provides the optimal policy for the arrival of data users (event E_2), the departure of voice and data users (events E_3, E_4 and E_5). Since this is an unconstrained problem, we set $\beta = 0$. The procedure Threshold−Policy−Search of

the MDP-based association algorithm computes the optimal policy for the arrival of voice users by solving an unconstrained MDP problem. The proposed procedure is motivated from well-known policy iteration algorithm [12]. However, the threshold nature of optimal policy (as derived in Lemma 5 and 6) is exploited which offers a significant reduction in computational complexity over policy iteration. We initially choose an arbitrary threshold th_0. In the POLICY–EVALUATION procedure, the system of value function equations is solved to determine the value function vector corresponding to thresholds th_n, $th_n + 1$ and $th_n - 1$. In the POLICY–IMPROVEMENT procedure, if the policy corresponding to the threshold $th_n + 1$ (or $th_n - 1$) improves the value function vector with at least one strict inequality (denoted by $\succeq$), the threshold is updated to $th_n + 1$ (or $th_n - 1$). In other words, compared to th_n, the policy corresponding to the threshold th_{n+1} has lesser number of states where the action is suboptimal. The procedure stops when no further improvement in value function vector is possible. Since in every iteration the policy improves and remains within the set of threshold policies, this algorithm converges to the optimal threshold policy. The calculated thresholds for the association of voice/data users are made available to the centralized controller connected to both the LTE BS and the WiFi AP. Since the centralized controller has an overall view of the whole system, the information regarding the numbers of active voice and data users in LTE and WiFi, are available to it. Whenever there is an arrival or a departure, the controller initiates the procedure POLICY–IMPL, as described in the MDP-based association algorithm. This procedure determines the state of the system based on the number of active users in LTE and WiFi networks and then chooses an appropriate action based on the corresponding thresholds.

B. CMDP-based Association Algorithm

We describe the CMDP-based algorithm which addresses the issue of high blocking probability of voice users, which may be encountered in the MDP-based association algorithm. The details of the algorithm are described in Algorithm 2. Apart from the same set of input parameters as required by the MDP-based association algorithm, the CMDP-based association algorithm requires $B_{\max}$ as an additional parameter to specify the constraint on the blocking probability of voice users. The procedure CALC–OPT–POLICY in the CMDP-based association algorithm computes the randomized optimal policy for the considered CMDP problem. First, the optimal policy is determined for a fixed value of β using same methodologies as in the MDP-based association algorithm. Then the value of β is updated until the difference of B^{π_β} and $B_{\max}$ becomes less than θ, where θ is a very small positive real number. All other procedures are similar to the procedures described in the MDP-based association algorithm.

C. Complexity Analysis

In this subsection, we investigate the gain in computational complexities of the proposed algorithms in comparison to traditional policy iteration algorithm [12], which performs the policy search over the entire policy space to determine the optimal policy. As discussed before, the computation of optimal policies for data user arrival, voice user departure, and data user departure is equivalent to the calculation of k_{th}

Algorithm 2 Network-Initiated CMDP-based Association Algorithm.

Input: $\lambda_v, \lambda_d, \mu_v, \mu_d, R_{L,V}, R_{L,D}, R_{W,D}(.), B_{\max}$.
1: Compute threshold k_{th} for the association of data users.
2: **procedure** CALC–OPT–POLICY
3:     Initialize β.
4:     **while** $|B^{\pi_\beta} - B_{\max}| > \theta$ **do**
5:         **procedure** THRESHOLD–POLICY–SEARCH
6:             See Algorithm 1.
7:         **end procedure**
8:         Update β using Equation (4).
9:     **end while**
10: **end procedure**
Output: Randomized optimal policy.
11: Compute thresholds $va_c(j,k)$ and $va_{lc}(j,k)$ for association of voice users for $(i + j) = C$ and $(i + j) < C$, respectively.
12: **procedure** POLICY–IMPL
13:     As discussed in Algorithm 1.
14: **end procedure**

---

with a computational complexity of $O(1)$. Now, we derive the computational complexity in obtaining the optimal policy for voice user arrival. In the MDP-based association algorithm, we evaluate three policies by solving a set of equations which can be solved using Gaussian elimination with associated complexity $O(|\mathcal{S}|^3)$, where $|\mathcal{S}|$ is the dimensionality of the state space. As described in section II, a state (i, j, k) in the state space $\mathcal{S}$ has the constraints $(i+j) \leq C$ and $k \leq W$. Therefore, the dimensionality of state space is $|\mathcal{S}| = O(C^2 W)$. Hence, each iteration of the policy evaluation can be performed in polynomial time as a function of C and W. In the policy improvement phase, the threshold structure of the optimal policies for the service of data user and voice user arrival are exploited to reduce the number of iterations in comparison to policy iteration algorithm. Without the knowledge of the threshold properties, the number of feasible policies for voice user arrival is $O(|\mathcal{A}|^{|\mathcal{S}|})$, where $|\mathcal{A}|$ is the size of the action space. To analyze how the threshold nature of optimal policy helps in reducing the number of feasible policies, we consider the following cases.

1) $0 \leq k < k_{th}$: As derived in Lemma 1, $k < k_{th} \implies j = 0$. According to Lemma 5, there exists a threshold on the value of i, where the optimal action changes to A_1. Since for every value of k, the threshold of blocking of voice users can be placed anywhere on the line $j = 0$, the number of policies in this region is equal to $C^{k_{th}}$.

2) $k = k_{th}$: Using Lemma 1 and 2, $k = k_{th} \implies j \geq 0$. For every value of j, the threshold of blocking of voice users (see Lemma 5) can be placed in $(C - j)$ different ways. Thus, the total no of policies is $O(C!)$.

3) $W \geq k > k_{th}$: As established in the Lemma 3, $k > k_{th} \Longrightarrow (i + j) = C$. According to Lemma 6, there exists a threshold on the value of i, where the optimal action changes from A_4 to A_1. Thus, for every value of k, the threshold can be placed in C different ways. Thus, the total number of policies in this region is $C^{(W-k_{th})}$.

Therefore, the total number of feasible policies is $O(C!C^W)$, which is the number of iterations in the MDP-based as-

sociation algorithm in the worst case. This brings down the worst case computational complexity of the MDP-based association algorithm from the computational complexity in policy iteration, which is $O(|\mathcal{A}|^{C^2W})$, to $O(C!C^W)$. Thus, the knowledge of threshold properties helps in achieving faster convergence, in comparison to policy iteration. In addition, contrary to policy iteration, the policy improvement phase consists of a comparison of value functions, which reduces the computational complexity from $O(|\mathcal{S}|^2|\mathcal{A}|)$ to $O(|\mathcal{S}|)$.

The computational complexity of the the CMDP-based association algorithm for a single iteration of β is equal to the computational complexity of the MDP-based algorithm.

The threshold property also reduces the storage complexity of the policy. Without the knowledge of threshold nature of optimal policy, number of bits required to store the optimal policy is $O(C^2W)$. By virtue of the threshold properties, we need to store the value of a single threshold in the case of data users and $(k_{th} + C + (W - k_{th}))$ thresholds in the case of voice users, which corresponds to storing only $(C + W)$ bits (according to the analysis presented above), which is a considerable reduction in storage complexity as well.

VI. NUMERICAL AND SIMULATION RESULTS

In this section, the algorithms proposed in the last section are implemented in ns-3 to observe the performance of the proposed algorithms. Performance of the proposed algorithms in terms of blocking probability of voice users and the total system throughput is compared to the performance of on-the-spot WiFi offloading algorithm [14]. In this algorithm [14], data user chooses LTE, only when there is no WiFi coverage. Therefore, in the considered system model, with on-the-spot offloading, data users always get associated with WiFi until WiFi capacity is exhausted. Voice users always get associated with LTE BS, and when LTE capacity is full, they are blocked. We also compare the performance of the proposed algorithms with the LTE-preferred scheme. In this scheme, voice and data users are associated with LTE until LTE reaches its capacity. When LTE reaches its capacity, and a voice user arrives, one existing LTE data user is offloaded to WiFi unless WiFi system also reaches its capacity. Otherwise, the incoming voice user is blocked. Additionally, when a voice user from LTE departs, if there is more than or equal to one data user in WiFi, an existing WiFi data user is moved to LTE.

A. Simulation Model and Evaluation Procedure

The simulated network model consists of a 3GPP LTE BS and an IEEE 802.11g WiFi AP. All users are taken to be stationary. The AP is approximately 50 m away from the LTE BS, and data users are distributed uniformly within 30 m radius of the WiFi AP. The WiFi AP is assumed to be deployed by the same cellular operator and hence trusted from the point of view of interworking. LTE and WiFi network parameters used in the simulation, as summarized in Table III and IV, are based on 3GPP models [39]-[40] and saturation throughput [11] 802.11g WiFi model. Propagation delay in WiFi network is assumed to be negligible. We consider CBR traffic for voice and data users in LTE. The generation of a fixed rate uplink flow is implemented in ns-3 using an application developed by us, which works similar to the ON/OFF application. This ap-

plication creates sockets between the sender and the receiver, and fixed sized packets are transmitted from the sender to the receiver at a constant bit rate.

Table III: LTE Network Model.

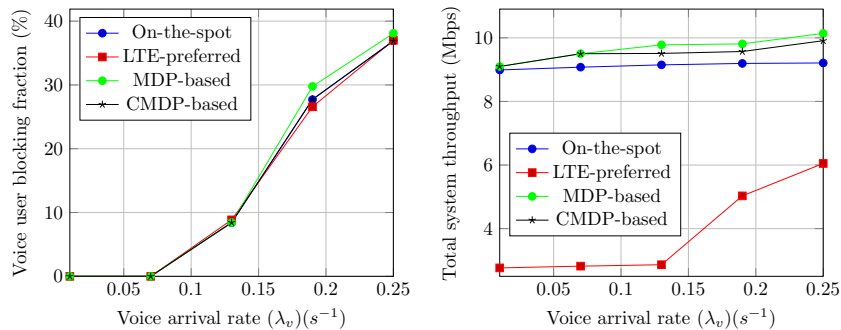| Parameter	Value
Maximum voice capacity	10 users
Maximum data capacity	10 users
Voice bit rate of a single user	20 kbps
Data bit rate of a single user	5 Mbps
Voice packet payload	50 bits
Data packet payload	600 bits
Tx power for BS and MS	46 dBm and 23 dBm
Noise figure for BS and MS	5 dB and 9 dB
Antenna height for BS and MS	32 m and 1.5 m
Antenna parameter for BS and MS	Isotropic antenna
Path loss	$128.1 + 37.6\log(R)$, R in kms

Table IV: WiFi Network Model.

Parameter	Value
Channel bit rate	54 Mbps
UDP header	224 bits
Packet payload	1500 bytes
Slot duration	20 μs
Short inter-frame space (SIFS)	10 μs
Distributed Coordination Function IFS (DIFS)	50 μs
Minimum acceptable per-user throughput	3.5 Mbps
Tx power for AP	23 dBm
Noise figure for AP	4 dB
Antenna height for AP	2.5 m
Antenna parameter	Isotropic antenna

B. Voice User Arrival Rate Variation

1) Voice User Blocking Probability Performance

Fig. 3a illustrates the variation of voice user blocking percentage of on-the-spot offloading [14], LTE-preferred, the MDP-based and CMDP-based association algorithms as a function of λ_v. In on-the-spot offloading, voice users are blocked when LTE reaches the capacity. When λ_v is small, the voice user blocking probability is small. However, as λ_v increases, the probability of approaching the LTE capacity and hence the voice user blocking probability increases. The voice user blocking probability in the MDP-based association algorithm is small when λ_v is small. However, as λ_v increases, voice user blocking probability values become marginally higher than the corresponding values for on-the-spot offloading. The MDP-based association algorithm may introduce blocking of voice users even when LTE has not reached its capacity, i.e., for states with $(i + j) < C$. Voice users have very less contribution to the total system throughput. Hence, voice users are blocked to save resources for data users which contribute significantly to the total system throughput. However, in the CMDP-based association algorithm, the number of states with proactive blocking (blocking when $(i+j) < C$) is reduced due to a constraint on the voice user blocking probability. Additionally, when i is small, the optimal action in states with $(i + j) = C$ becomes A_4 (accept voice user in LTE and data offload to WiFi). Voice user blocking probability contribution comes mainly from the states with $(i + j) = C$, where i is large (say states $(C, 0, 0), (C-1, 1, 0)$ etc.). Since a major fraction of voice user blocking occurs when $(i+j) = C$ and i is large, the system becomes analogous to the on-the-spot offloading. Hence, the voice user blocking probability performance of the CMDP-based association algorithm is almost similar to on-the-spot offloading algorithm [14]. Both on-the-spot offloading and LTE-preferred scheme block voice users

(a) Voice user blocking percentage vs. λ_v ($\lambda_d = 1/20, \mu_v = 1/60$ and $\mu_d = 1/10$).

(b) Total system throughput vs. λ_v ($\lambda_d = 1/20, \mu_v = 1/60$ and $\mu_d = 1/10$).

Figure 3: Plot of blocking fraction of voice users and total system throughput for different algorithms under varying λ_v.

only when LTE system is full with only voice users. Therefore, blocking probability performance of these two schemes are similar.

2) Total System Throughput Performance

Total system throughput performance comparison of different algorithms is illustrated in Fig. 3b. In on-the-spot offloading, the average number of voice users in LTE increases with λ_v, while the average number of data users in WiFi remains constant. Thus, the total system throughput increases with λ_v. For the MDP-based association algorithm, with an increase in λ_v, the blocking probability of voice users increases. Therefore, the fraction of voice users in the system decreases, and the total system throughput increases. Besides, the MDP-based association algorithm performs a significant amount of load balancing under A_4 (accept voice user in LTE with data user offload to WiFi) and A_5 (move data user to the RAT from where a user has departed). With higher λ_v, load balancing actions are chosen more frequently. Thus, with higher λ_v, the MDP-based association algorithm exhibits greater improvement over on-the-spot offloading algorithm. The improvement in total system throughput varies from 1.22% (for $\lambda_v = 0.01$) to 10.32% (for $\lambda_v = 0.25$). In Fig. 3b, we observe that the CMDP-based association algorithm also performs better than on-the-spot offloading. However, due to a constraint on the voice user blocking probability, the performance improvement is lower than the MDP-based association algorithm. For lower values of λ_v ($\lambda_v = 0.01, 0.07$), the total system throughput of the CMDP-based association algorithm is same as that of the MDP-based association algorithm as the optimal policy for the CMDP is same as that of the unconstrained MDP. On-the-spot offloading algorithm blocks voice users only when LTE reaches capacity. Typically, in the CMDP-based association algorithm also, voice user blocking occurs when the LTE is full with a large number of voice users. However, due to load balancing of data users, the CMDP-based association algorithm outperforms the on-the-spot offloading algorithm. With $\lambda_v = 0.01$, the improvement in total system throughput is only 1.22% and with $\lambda_v = 0.25$, it becomes 7.60%. LTE-preferred scheme associates both voice and data users to LTE. As λ_v increases, the probability that LTE reaches its capacity, increases. Therefore, to accommodate incoming voice users, existing LTE data users are offloaded to WiFi, which increases the total system throughput when WiFi load is low. However,

both the MDP-based and CMDP-based association algorithms outperform LTE-preferred scheme.

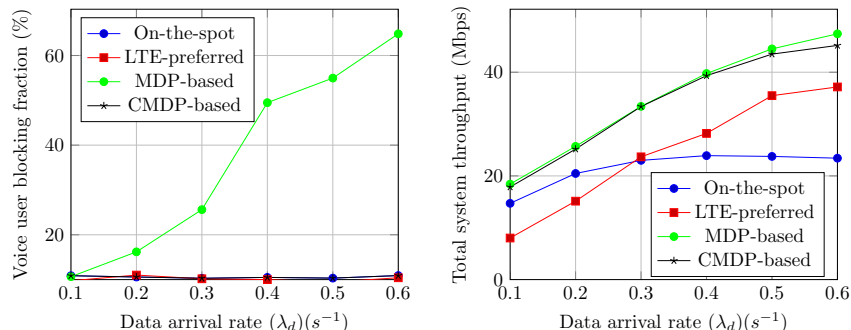### C. Data User Arrival Rate Variation

1) Voice User Blocking Probability Performance

In Fig. 4a, for on-the-spot offloading, voice and data users are accepted in LTE and WiFi, respectively. Consequently, changes in λ_d do not affect the blocking probability performance of voice users in LTE. In the case of the MDP-based association algorithm, increase in λ_d associates more number of data users with LTE, since the optimal policy for data users is to associate with LTE after the number of WiFi data users crosses a certain threshold. Therefore, the number of free LTE resources for voice users reduces, eventually increasing the blocking probability of voice users. The voice user blocking probability of the MDP-based association algorithm is worse than that of on-the-spot offloading and increases with λ_d. The blocking probability performance of the CMDP-based association algorithm and LTE-preferred scheme are similar to that of the on-the-spot offloading. Since usually the voice users are blocked in the states where the only feasible action is blocking (say state $(C, 0, 0)$), the decision epochs where voice users are blocked are almost same as that of the on-the-spot offloading and LTE-preferred scheme.

2) Total System Throughput Performance

In Fig. 4b, total system throughput for different algorithms are plotted as a function of λ_d.

In on-the-spot offloading, with an increase in λ_d, the number of WiFi data users increases, and this increases the total system throughput. However, for high λ_d, the effect of contention among data users reduces the rate of increment of the total system throughput. In the MDP-based association algorithm, as λ_d increases, more number of data users are served using LTE. Since the throughput contribution of data users is more than voice users, the blocking probability of voice users increases with λ_d. Thus, the fraction of voice users in the system reduces, effectively causing more improvement in the total system throughput. When $\lambda_d = 0.1$, the improvement in system metric is 25.22%, whereas for $\lambda_d = 0.6$, the system metric almost doubles. In Fig. 4b, the total system throughput values for the CMDP-based association algorithm are smaller than the corresponding values for the MDP-based association algorithm. The reduction in blocking probability of voice users comes at a price of the reduction in the total

(a) Voice user blocking percentage vs. λ_d ($\lambda_v = 1/6, \mu_v = 1/60$ and $\mu_d = 1/10$).

(b) Total system throughput vs. λ_d ($\lambda_v = 1/6, \mu_v = 1/60$ and $\mu_d = 1/10$).

Figure 4: Plot of blocking fraction of voice users and total system throughput for different algorithms under varying λ_d.

system throughput. Still, due to optimal association and load balancing decisions, the CMDP-based association algorithm reduces the effect of contention among data users in WiFi and hence performs better than on-the-spot offloading algorithm [14]. For example, with $\lambda_d = 0.1$, the improvement in system metric is about 22.96% and with $\lambda_v = 0.6$, it becomes almost 93%. Similar to Fig. 3b, in case of the LTE-preferred scheme, as λ_d increases, more data users are offloaded to WiFi, resulting in an improvement in total system throughput. Since the effect of contention in the LTE-preferred scheme is lesser than that of the on-the-spot offloading, LTE-preferred scheme outperforms the on-the-spot offloading under large λ_d. However, the performance of the LTE-preferred scheme is worse than both the MDP-based and the CMDP-based association algorithms.

VII. Conclusion

In this paper, we have formulated the optimal association problem in an LTE-WiFi HetNet as an MDP problem with an objective of maximizing the total system throughput. Constrained MDP formulation has also been presented, where maximizing the total system throughput is subject to a constraint on the blocking probability of voice users. Threshold structures on the association of voice and data users have been derived. Based on the structure of the optimal policies, we have proposed two algorithms for the association and offloading of voice/data users in an LTE-WiFi HetNet. Analysis indicates that the knowledge of threshold structures reduces the number of feasible policies and hence reduces the computational complexity considerably (from $O(|\mathcal{A}|^{C^2 W})$ to $O(C!C^W)$) in comparison to policy iteration. Simulation results demonstrate that although the voice user blocking probability performance of the MDP-based association algorithm is worse than that of on-the-spot offloading [14], the CMDP-based association algorithm performs as good as on-the-spot offloading and LTE-preferred scheme. Moreover, the proposed algorithms perform better than both the on-the-spot offloading algorithm and LTE-preferred scheme in improving the total system throughput. In future, the considered framework can be extended to consider the channel state between LTE BS/WiFi AP and users such that channel-aware association and offloading decisions can be taken.
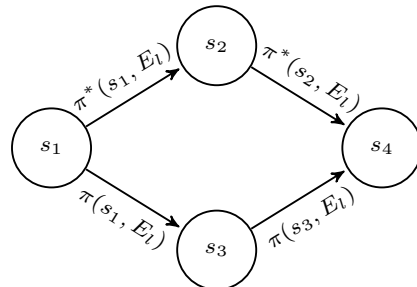
Appendix A
Proof of Lemma 1

Figure 5: Sample path under different policies.

Since the decisions of association and offloading are involved during the arrival and departure of users, proving this lemma is equivalent to proving the following statements.

(a) A_3 (Accept in WiFi) is optimal when there are less than k_{th} data users in the system, and a data user arrives.

(b) A_1 (Do nothing) is optimal when there are less than or equal to k_{th} data users in the system, and a voice user from LTE departs.

(c) A_1 (Do nothing) is optimal when there are less than or equal to k_{th} data users in the system, and a data user from WiFi departs.

We prove the above statements by sample path argument. Suppose the system starts at time $t = 0$.

Proof of (a): We consider the scenario when the system is in the state $s_1 = (i, 0, 0)$, when a data user arrival occurs (after a time t_1, say). Assume that the optimal policy π^* does not associate this incoming data user with WiFi. Therefore, the optimal action must be A_2 (Accept in LTE). As the optimal policy is π^*, we have $V^{\pi^*}(s) \geq V^{\hat{\pi}}(s)$, $\forall \hat{\pi} \in \prod$ and $\forall s \in \mathcal{S}$, where $\prod$ is the set of all policies. Let us consider another policy π (non-stationary in general) which takes A_3 in state $s_1 = (i, 0, 0)$. As illustrated in Fig. 5, let us assume that starting from the state s_1 and following the policy π^* and π, the system reaches the state $s_2 = (i, 1, 0)$ and $s_3 = (i, 0, 1)$, respectively. The inter-arrival times and service times are same for both the sample paths as we have considered a Markovian system. Assume that from the state s_2, based on the next event (after a time t_2) and the chosen action, the system makes a transition to the state s_4 according to the policy π^*. Before

reaching the state s_2, the sample path followed by the policy π^* has one less WiFi data user and one more LTE data user than that of the policy π before it reaches the state s_3. Suppose, the policy π is such that in state s_3, it takes the same action as that of policy π^* and additionally offloads one data user from WiFi to LTE. Evidently, sample path followed by both the policies end up in the same state s_4. We construct π in such a manner that from the state s_4 onwards, both the policies take up same actions and follow the same sample path. Therefore, the difference of value functions of the state s_1 under the policy π^* and π is

$$V^{\pi^*}(s_1) - V^{\pi}(s_1) = R_{L,D} - R_{W,D}(1).$$

Since $R_{L,D} < \tilde{R}_{W,D}(k), \forall k < k_{th}$ and $\tilde{R}_{W,D}(1) = R_{W,D}(1)$, we have, $V^{\pi^*}(s_1) < V^{\pi}(s_1)$. Clearly, this contradicts the original claim that π^* is an optimal policy. Since the Markov chains induced by different policies are recurrent, the state $(i,0,0)$ is visited infinitely often and each time choice of A_3 upon a data user arrival provides more reward than action A_2. Therefore, when there is no data user in the system, and one data user arrives, A_3 is optimal. In a similar manner, it can be proved that A_3 is optimal when a data user arrives and the system is in state $(i,0,k)$, where $k < k_{th}$.

Proof of (b) and (c): These can be proved using a similar sample path argument.

APPENDIX B
PROOF OF LEMMA 2

Similar to Lemma 1, proving this lemma is equivalent to proving the following statements.
(a) A_2 (Accept in LTE) is optimal when there are more than or equal to k_{th} data users in the system, and one data user arrives.
(b) A_1 (Do nothing) is optimal when there are more than k_{th} data users in the system, and a voice/data user from LTE departs.
(c) A_5 (Data offload to a RAT from where a user has departed) is optimal when there are more than k_{th} data users in the system, and a data user from WiFi departs.
Proof of (a): From Lemma 1, we have, $(j + k) \leq k_{th} \implies j = 0$. We consider the scenario when the system is in the state $(i, 0, k_{th})$, when a data user arrival occurs. Assume that the optimal policy π^* does not associate this incoming data user with LTE. Consequently, the optimal action must be A_3. As the optimal policy is π^*, we have $V^{\pi^*}(s) \geq V^{\hat{\pi}}(s) \, \forall \hat{\pi} \in \prod$ in every state s. Let us consider another policy π which chooses A_2 in state $(i, 0, k_{th})$. As illustrated in Fig. 5, starting from the state $(i, 0, k_{th})$ and following the policy π^* and π, the system reaches the states s_2 and s_3, respectively. From state s_2, based on an event, the system reaches the state s_4. Suppose, in the state s_3, the action followed by policy π is such that it chooses the same action as that of policy π^* and additionally offloads one data user from LTE to WiFi. Clearly, path followed by both the policies end up in the same state s_4. We construct π in such a way that from the state s_4 onwards, both of them follow the same path. Similar to the previous lemma, the difference of value functions under the policy π^*

and π is

$$V^{\pi^*}(s_1) - V^{\pi}(s_1) = \big((k+1)R_{W,D}(k+1) - kR_{W,D}(k) - R_{L,D}\big).$$

Since $R_{L,D} \geq \tilde{R}_{W,D}(k), \forall k \geq k_{th}$, we have, $V^{\pi^*}(s_1) < V^{\pi}(s_1)$. Clearly, this contradicts the original claim that π^* is an optimal policy. Thus, A_2 is optimal when there are k_{th} data users in WiFi, and one data user arrives. The same result can be extended for the case when there are k_{th} data users in WiFi, more than or equal to one data user in LTE, and one data user arrives.

Statements (b) and (c) can be proved in a similar way.

APPENDIX C
PROOF OF LEMMA 5

To prove this lemma, we consider two cases, (1) $k \geq k_{th}$ and (2) $k < k_{th}$. We prove the required for the first case. Proof of the second case follows in a similar manner. From Lemma 4, we know that for $k \geq k_{th}$, A_2 is better than A_4. Thus, for $k \geq k_{th}$, the choice is between A_1 and A_2. To prove this lemma, we first prove that the value function $V(i, j, k)$ is concave in i. In Lemma 1 and 2, we have already derived the structure of the optimal policy for data user arrival and departure of voice and data users. Now, for $k \geq k_{th}$, with the aid of this, the optimality equation is as follows.

$$\begin{aligned}
V(i,j,k) &= \lambda_v \delta \max\{f(i,j,k) - \beta + V(i,j,k), f(i+1,j,k) \\
&\quad + V(i+1,j,k)\} + \lambda_d \delta\big(f(i,j+1,k) + V(i,j+1,k)\big) \\
&\quad + i\mu_v \delta\big(f(i-1,j+1,k-1) + V(i-1,j+1,k-1)\big) \\
&\quad + j\mu_d \delta\big(f(i,j,k-1) + V(i,j,k-1)\big) \\
&\quad + k\mu_d \delta\big(f(i,j,k-1) + V(i,j,k-1)\big) \\
&\quad + \big(1 - v(i,j,k)\big)V(i,j,k).
\end{aligned} \tag{6}$$

Let the components in Equation (6) be denoted by $V^1(i,j,k), V^2(i,j,k), V^3(i,j,k), V^4(i,j,k), V^5(i,j,k)$ and $V^6(i,j,k)$, respectively. We prove the concavity of $V(i,j,k)$ component-wise. Start the VIA with $V_0(i,j,k) = 0$. Hence, $V_0(i,j,k)$ is concave in i. Let us assume that $V_{1,n}(i,j,k) = \max\{f(i,j,k) - \beta + V_{n-1}(i,j,k), f(i+1,j,k) + V_{n-1}(i+1,j,k)\}$. Equivalently, $V_{1,n}(i,j,k) = \max\{-\beta + V_{n-1}(i,j,k), R_{L,V} + V_{n-1}(i+1,j,k)\}$. Let us define the function $V_{1,n}(i,j,k,a)$ as follows.

$$V_{1,n}(i,j,k,a) = \begin{cases} -\beta + V_{n-1}(i,j,k), & a = A_1, \\ R_{L,V} + V_{n-1}(i+1,j,k), & a = A_2. \end{cases}$$

By definition,

$$V_{1,n}(i,j,k) = \max_{a \in \{A_1, A_2\}} V_{1,n}(i,j,k,a).$$

Thus, we have,

$$V^1(i,j,k) = \lim_{n \to \infty} V_{1,n}(i,j,k).$$

Let us define $D_i V(i,j,k,a) = V(i+1,j,k,a) - V(i,j,k,a)$.

$$D_i V_{1,n}(i,j,k,a) = \begin{cases} D_i V_{n-1}(i,j,k), & a = A_1, \\ D_i V_{n-1}(i+1,j,k), & a = A_2. \end{cases}$$

$$D_{ii} V_{1,n}(i,j,k,a) = \begin{cases} D_{ii} V_{n-1}(i,j,k), & a = A_1, \\ D_{ii} V_{n-1}(i+1,j,k), & a = A_2. \end{cases}$$

Since $V_{n-1}(i,j,k)$ is concave in i, $V_{1,n}(i,j,k,a)$ is concave in i.

Now, we need to prove that $V_{1,n}(i,j,k)$ is concave in i. In other words, we need to prove that $V_{1,n}(i+2,j,k) + V_{1,n}(i,j,k) \leq 2V_{1,n}(i+1,j,k)$. Let us assume that $a_1 \in \{A_1, A_2\}$ and $a_2 \in \{A_1, A_2\}$ are the maximizing actions in states $(i+2,j,k)$ and (i,j,k), respectively. Therefore,

$$2V_{1,n}(i+1,j,k) \geq V_{1,n}(i+1,j,k,a_1) + V_{1,n}(i+1,j,k,a_2)$$
$$= V_{1,n}(i+2,j,k,a_1) + V_{1,n}(i,j,k,a_2)$$
$$- D_i V_{1,n}(i+1,j,k,a_1) + D_i V_{1,n}(i,j,k,a_2).$$

Let us take $X = D_i V_{1,n}(i,j,k,a_2) - D_i V_{1,n}(i+1,j,k,a_1)$. To prove that $V_{1,n}(i,j,k)$ is concave in i, we need to prove that $X \geq 0$. There are four cases as described below.
Case 1 : $a_1 = a_2 = A_1$,

$$X = D_i V_{n-1}(i,j,k) - D_i V_{n-1}(i+1,j,k)$$
$$= -D_{ii} V_{n-1}(i,j,k) \geq 0.$$

Case 2 : $a_1 = A_1, a_2 = A_2$,

$$X = D_i V_{n-1}(i+1,j,k) - D_i V_{n-1}(i+1,j,k) = 0.$$

Case 3 : $a_2 = a_2 = A_2$,

$$X = D_i V_{n-1}(i+1,j,k) - D_i V_{n-1}(i+2,j,k)$$
$$= -D_{ii} V_{n-1}(i+1,j,k) \geq 0.$$

Case 4 : $a_1 = A_2, a_2 = A_1$,

$$X = D_i V_{n-1}(i,j,k) - D_i V_{n-1}(i+2,j,k)$$
$$= -D_{ii} V_{n-1}(i,j,k) - D_{ii} V_{n-1}(i+1,j,k) \geq 0.$$

Thus, it is proved that $V_{1,n}(i,j,k)$ is concave in i. Since this holds for every n and every β, $V^1(i,j,k)$ is concave in i.

Similarly, in the case of the second component, let $V_{2,n}(i,j,k) = f(i+1,j,k) + V_{n-1}(i,j+1,k)$.

Thus, $D_{ii} V_{2,n}(i,j,k) = D_{ii} V_{n-1}(i,j+1,k)$. Therefore, $V_{2,n}(i,j,k)$ is concave in i. Similarly, other components also can be proved to be concave in i. Therefore, $V(i,j,k)$ is concave in i.

Let us define $x(i,j,k) = -\beta - R_{L,V}$. In order to prove this lemma, we know that if state (i,j,k) is blocking, then $V(i+1,j,k) - V(i,j,k) \leq x(i,j,k)$. Due to concavity of $V(i,j,k)$, $V(i+2,j,k) - V(i+1,j,k) \leq V(i+1,j,k) - V(i,j,k)$. Now, $x(i,j,k) = x(i+1,j,k)$. As a consequence, $V(i+2,j,k) - V(i+1,j,k) \leq x(i+1,j,k)$. Thus, it is proved that if state (i,j,k) is blocking, then the state $(i+1,j,k)$ is also blocking.

To prove that if state (i,j,k) is blocking, then the state $(i,j+1,k)$ is also blocking, we first need to prove that the value function is submodular in (i,j). In other words, we need to prove that $V_n(i+1,j,k) + V_n(i,j+1,k) \geq V_n(i,j,k) + V_n(i+1,j+1,k)$. Similar to the previous proof, we prove the above statement component-wise. Let us assume that a_1 and a_2 are the maximizing actions in states (i,j,k) and $(i+1,j+1,k)$, respectively. Start the VIA with $V_0(i,j,k) = 0$.

Therefore, $V_0(i,j,k)$ is submodular in (i,j). In other words, $D_{ij} V_0(i,j,k) \leq 0$. We have,

$$V_{1,n}(i+1,j,k) + V_{1,n}(i,j+1,k)$$
$$\geq V_{1,n}(i+1,j,k,a_1) + V_{1,n}(i,j+1,k,a_2)$$
$$= V_{1,n}(i,j,k,a_1) + V_{1,n}(i+1,j+1,k,a_2)$$
$$+ D_i V_{1,n}(i,j,k,a_1) - D_i V_{1,n}(i,j+1,k,a_2).$$

Now, we consider four possible cases.
Case 1 : $a_1 = a_2 = A_1$,

$$D_i V_{1,n}(i,j,k,a_1) - D_i V_{1,n}(i,j+1,k,a_2)$$
$$= D_i V_{n-1}(i,j,k) - D_i V_{n-1}(i,j+1,k)$$
$$= -D_{ij} V_{n-1}(i,j,k) \geq 0.$$

Case 2 : $a_1 = a_2 = A_2$,

$$D_i V_{1,n}(i,j,k,a_1) - D_i V_{1,n}(i,j+1,k,a_2)$$
$$= D_i V_{n-1}(i+1,j,k) - D_i V_{n-1}(i+1,j+1,k)$$
$$= -D_{ij} V_{n-1}(i+1,j,k) \geq 0.$$

Case 3 : $a_1 = A_1, a_2 = A_2$,

$$D_i V_{1,n}(i,j,k,a_1) - D_i V_{1,n}(i,j+1,k,a_2)$$
$$= D_i V_{n-1}(i,j,k) - D_i V_{n-1}(i+1,j+1,k)$$
$$= D_i V_{n-1}(i,j,k) - D_i V_{n-1}(i,j+1,k)$$
$$+ D_i V_{n-1}(i,j+1,k) - D_i V_{n-1}(i+1,j+1,k)$$
$$= -D_{ij} V_{n-1}(i,j,k) - D_{ii} V_{n-1}(i,j+1,k) \geq 0.$$

Case 4 : $a_1 = A_2, a_2 = A_1$,

$$V_{1,n}(i+1,j,k) + V_{1,n}(i,j+1,k)$$
$$\geq V_{1,n}(i+1,j,k,a_2) + V_{1,n}(i,j+1,k,a_1)$$
$$= -\beta + V_{n-1}(i+1,j,k) + R_{L,V} + V_{n-1}(i+1,j+1,k)$$
$$= V_{1,n}(i,j,k,2) + V_{1,n}(i+1,j+1,k,1)$$
$$= V_{1,n}(i,j,k) + V_{1,n}(i+1,j+1,k).$$

Thus, it is proved that $V_{1,n}(i,j,k)$ is submodular in (i,j).

Similarly, in the case of the second component, we have, $V_{2,n}(i,j,k) = f(i,j+1,k) + V_n(i,j+1,k)$. Therefore, we have, $D_{ij} V_{2,n}(i,j,k) = D_{ij} V_n(i,j+1,k) \leq 0$. Similarly, other components also can be proved to be submodular in (i,j). Therefore, the value function is submodular in (i,j).

Now, if state (i,j,k) is blocking then we have, $V(i+1,j,k) - V(i,j,k) \leq x(i,j,k)$. Again, we have, $x(i,j,k) = x(i,j+1,k)$. Due to submodularity, we have $V(i+1,j+1,k) - V(i,j+1,k) \leq V(i+1,j,k) - V(i,j,k) \leq x(i,j,k) = x(i,j+1,k)$. Thus, in the case of voice arrival, if A_1 is optimal in state (i,j,k), then in state $(i,j+1,k)$ also A_1 is optimal.

Proof of (ii) follows directly from the proof of part (i).

APPENDIX D
PROOF OF LEMMA 6

To prove this lemma, we consider two cases, (1) $k \geq k_{th}$ and (2) $k < k_{th}$. We demonstrate the proof of the lemma for the first case. Proof of the second case follows in similar manner. To prove this lemma, we first need to prove that for $(i+j) = C$, the difference of value functions $V(i+1,j-$

$1, k+1) - V(i, j, k)$ is decreasing in i. For $(i + j) = C$ and $k \geq k_{th}$, the optimality equation can be described as

$$
\begin{aligned}
V(i, j, k) =\ & \lambda_v \delta \max\{f(i, j, k) - \beta + V(i, j, k), \\
& f(i+1, j-1, k+1) + V(i+1, j-1, k+1)\} \\
& + \lambda_d \delta \big(f(i, j, k+1) + V(i, j, k+1)\big) \\
& + i\mu_v \delta \big(f(i-1, j+1, k-1) + V(i-1, j+1, k-1)\big) \\
& + j\mu_d \delta \big(f(i, j, k-1) + V(i, j, k-1)\big) \\
& + k\mu_d \delta \big(f(i, j, k-1) + V(i, j, k-1)\big) \\
& + \big(1 - v(i, j, k)\big) V(i, j, k).
\end{aligned}
$$

Let us assume that $V_{1,n}(i, j, k, a) = \max\{f(i, j, k) - \beta + V_{n-1}(i, j, k), f(i+1, j-1, k+1) + V_{n-1}(i+1, j-1, k+1)\}$. Equivalently, $V_{1,n}(i, j, k, a) = \max\{-\beta + V_{n-1}(i, j, k), R_{L,V} - R_{L,D} + \tilde{R}_{W,D}(k) + V_{n-1}(i+1, j-1, k+1)\}$.

In other words,

$$
V_{1,n}(i, j, k, a) = \begin{cases} -\beta + V_{n-1}(i, j, k), & a = A_1, \\ R_{L,V} - R_{L,D} + \tilde{R}_{W,D}(k) + \\ V_{n-1}(i+1, j-1, k+1), & a = A_4. \end{cases}
$$

We prove the above claim component-wise. Start the VIA with $V_0(i, j, k) = 0$. Therefore, $V_0(i+1, j-1, k+1) - V_0(i, j, k)$ is decreasing in i. We also have, $E_i F_i V_0(i, j, k) \leq 0$. Now, Let us define $E_i V(i, j, k, a) = V(i+1, j-1, k+1, a) - V(i, j, k, a)$ and $F_i V(i, j, k, a) = V(i+1, j-1, k, a) - V(i, j, k, a)$.

$$
E_i V_{1,n}(i, j, k, a) = \begin{cases} E_i V_{n-1}(i, j, k), & a = A_1, \\ E_i V_{n-1}(i+1, j-1, k+1), & a = A_4. \end{cases}
$$

Therefore, $E_i F_i V_{1,n}(i, j, k, a) \leq 0$.

$$
E_{ii} V_{1,n}(i, j, k, a) = \begin{cases} E_{ii} V_{n-1}(i, j, k), & a = A_1, \\ E_{ii} V_{n-1}(i+1, j-1, k+1), & a = A_4. \end{cases}
$$

Therefore, $V_{1,n}(i+1, j-1, k+1, a) - V_{1,n}(i, j, k, a)$ is decreasing in i.

Now, we need to prove that $V_{1,n}(i+1, j-1, k+1) - V_{1,n}(i, j, k)$ is decreasing in i. In other words, we need to prove that $V_{1,n}(i+2, j-2, k+1) + V_{1,n}(i, j, k) \leq V_{1,n}(i+1, j-1, k+1) + V_{1,n}(i+1, j-1, k)$. Let us assume that $a_1 \in \{A_1, A_4\}$ and $a_2 \in \{A_1, A_4\}$ are the maximizing actions in states $(i+2, j-2, k+1)$ and (i, j, k), respectively. Therefore,

$$
\begin{aligned}
& V_{1,n}(i+1, j-1, k+1) + V_{1,n}(i+1, j-1, k) \\
& \geq V_{1,n}(i+1, j-1, k+1, a_2) + V_{1,n}(i+1, j-1, k, a_1) \\
& = V_{1,n}(i, j, k, a_2) + V_{1,n}(i+2, j-2, k+1, a_1) \\
& - E_i V_{1,n}(i+1, j-1, k, a_1) + E_i V_{1,n}(i, j, k, a_2).
\end{aligned}
$$

Let us take $Y = E_i V_{1,n}(i, j, k, a_2) - E_i V_{1,n}(i+1, j-1, k, a_1)$. To prove that $V_{1,n}(i+1, j-1, k+1) - V_{1,n}(i, j, k)$ is decreasing in i, we need to prove that $Y \geq 0$. There are four cases as described below.

Case 1 : $a_1 = a_2 = A_1$,

$$
\begin{aligned}
Y &= E_i V_{n-1}(i, j, k) - E_i V_{n-1}(i+1, j-1, k) \\
&= -E_i F_i V_{n-1}(i, j, k) \geq 0.
\end{aligned}
$$

Case 2 : $a_1 = A_4, a_2 = A_1$,

$$
\begin{aligned}
Y &= E_i V_{n-1}(i, j, k) - E_i V_{n-1}(i+2, j-2, k+1) \\
&= -E_{ii} V_{n-1}(i, j, k) - E_i F_i(i+1, j-1, k+1) \geq 0.
\end{aligned}
$$

Case 3 : $a_2 = a_2 = A_4$,

$$
\begin{aligned}
Y &= E_i V_{n-1}(i+1, j-1, k+1) - E_i V_{n-1}(i+2, j-2, k+1) \\
&= -E_i F_i V_{n-1}(i+1, j-1, k+1) \geq 0.
\end{aligned}
$$

Case 4 : $a_1 = A_1, a_2 = A_4$,

$$
\begin{aligned}
& V_{1,n}(i+1, j-1, k+1) + V_{1,n}(i+1, j-1, k) \\
& \geq V_{1,n}(i+1, j-1, k+1, a_1) + V_{1,n}(i+1, j-1, k, a_2) \\
& = -\beta + V_{n-1}(i+1, j-1, k+1) + R_{L,V} - R_{L,D} + \\
& \tilde{R}_{W,D}(k) + V_{n-1}(i+2, j-2, k+1) \\
& = V_{1,n}(i+2, j-2, k+1, 1) + V_{1,n}(i, j, k, 4) \\
& = V_{1,n}(i+2, j-2, k+1) + V_{1,n}(i, j, k).
\end{aligned}
$$

Thus, it is proved that $V_{1,n}(i+1, j-1, k+1) - V_{1,n}(i, j, k)$ is decreasing in i. Since this holds for every n and every value of β, $V^1(i+1, j-1, k+1) - V^1(i, j, k)$ is decreasing in i.

Let $V_{2,n}(i, j, k) = f(i, j, k+1) + V_{n-1}(i, j, k+1)$. Thus, $E_{ii} V_{2,n}(i, j, k) = E_{ii} V_{n-1}(i, j, k+1)$. Therefore, $V_{2,n}(i+1, j-1, k+1) - V_{2,n}(i, j, k)$ is decreasing in i. Similarly, other components can be proved to be decreasing in i. Therefore, $V(i+1, j-1, k+1) - V(i, j, k)$ is decreasing in i.

Similar to Lemma 5, using this property it can be proved that if the optimal action for voice user arrival in state (i, j, k) is blocking, then the optimal action in state $(i+1, j-1, k)$ is also blocking.

Proof of (ii) follows directly from the proof of part (i).

ACKNOWLEDGMENT

REFERENCES

[1] F. Rebecchi, M. D. de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data Offloading Techniques in Cellular Networks: a Survey," IEEE Communications Surveys & Tutorials, Vol. 17, No. 2, pp. 580-603, Nov 2014.

[2] Y. He, M. Chen, B. Ge, and M. Guizani, "On WiFi offloading in heterogeneous networks: Various incentives and trade-off strategies," IEEE Communications Surveys & Tutorials, Vol. 18, No. 4, pp. 2345-2385, Apr 2016.

[3] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021, Mar 2017.

[4] IEEE 802.11-2012, Part 11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," Mar 2012.

[5] 3GPP TS 23.234 v12.0.0, "3GPP System to Wireless Local Area Network (WLAN) Interworking; System Description," Available online at http://www.3gpp.org/DynaReport/23234.htm, Sep 2014.

[6] M. Gerasimenko, N. Himayat, S.P. Yeh, S. Talwar, S. Andreev, and Y. Koucheryavy, "Characterizing Performance of Load-aware Network Selection in Multi-radio (WiFi/LTE) Heterogeneous Networks," in GLOBECOM Workshop, pp. 397-402, Dec 2013.

[7] F. Moety, M. Ibrahim, S. Lahoud, and K. Khawam, "Distributed Heuristic Algorithms for RAT Selection in Wireless Heterogeneous Networks," in proc. of WCNC, pp. 2220 - 2224, Apr 2012.

[8] A. Whittier, P. Kulkarni, F. Cao, and S. Armour, "Mobile data offloading addressing the service quality vs. resource utilisation dilemma," in proc. of PIMRC, pp. 1-6, Sep 2016.

[9] M. Yassin, M. Ibrahim, and S. Lahoud, "A Hybrid Approach for RAT Selection in Wireless Heterogeneous Networks," in proc. of ICCIT, pp. 290-294, Jun 2013.

[10] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT Selection Games in HetNets," in proc. of INFOCOM, pp. 998-1006, Apr 2013.

[11] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," IEEE Journal on Selected Areas in Communications, Vol. 18, No. 3, pp. 535-547, Mar 2000.

[12] M. L. Putterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley and Sons, 1994.

[13] http://code.nsnam.org/ns-3-dev/.

[14] F. Mehmeti and T. Spyropoulos, "Performance Analysis of "On-the-spot" Mobile Data Offloading," in proc. of GLOBECOM, pp. 1577 - 1583, Dec 2013.

[15] D. Pacheco-Paramo, V. Pla, V. Casares-Giner, and J. Martinez-Bauset, "Optimal Radio Access Technology Selection on Heterogeneous Networks," Physical Communication, Vol. 5, Issue 3, pp. 253-271, Sep 2012.

[16] A. Kumar and V. Kumar, "Optimal Association of Stations and APs in an IEEE 802.11 WLAN," in proc. of NCC, pp. 1-5, Jan 2005.

[17] S. Barmpounakis, A. Kaloxylos, P. Spapis, and N. Alonistioti, "Context-aware, user-driven, network-controlled RAT selection for 5G networks," Computer Networks, 113, pp. 124-47, 2017.

[18] C. Liu, M. Li, S.V. Hanly, and P. Whiting, "Joint downlink user association and interference management in two-tier HetNets with dynamic resource partitioning," IEEE Transactions on Vehicular Technology, Vol. 66, No. 2, pp. 1365-1378, Feb 2017.

[19] G.S. Kasbekar, P. Nuggehalli, and J. Kuri, "Online Client-AP Association in WLANs," in proc. of WiOpt, pp. 1-8, Apr 2006.

[20] D. Kumar, E. Altman, and J. Kelif, "Globally Optimal User-Network Association in an 802.11 WLAN & 3G UMTS Hybrid Cell," in proc. of ITC, pp. 1173-1187, Jun 2007.

[21] E. Khloussy, X. Gelabert, and Y. Jiang, "A Revenue-Maximizing Scheme for Radio Access Technology Selection in Heterogeneous Wireless Networks with User Profile Differentiation," Advances in Communication Networking, Springer, pp. 66-77, 2013.

[22] M. El Helou, M. Ibrahim, S. Lahoud, K. Khawam, D. Mezher, and B. Cousin, "A Network-assisted Approach for RAT Selection in Heterogeneous Cellular Networks," IEEE Journal on Selected Areas in Communications, Vol. 33, No. 6, pp. 1055–1067, Jun 2015.

[23] W. Song, Y. Cheng, and W. Zhuang, "Improving Voice and Data Services in Cellular/WLAN Integrated Networks by Admission Control," IEEE Transactions on Wireless Communications, Vol. 6, No. 11, pp. 4025-4037, Nov 2007.

[24] W. Song, H. Jiang, W. Zhuang, and A. Saleh, "Call Admission Control for Integrated Voice/data Services in Cellular/WLAN Interworking," in proc. of ICC, pp. 5480-5485, Jun 2006.

[25] W. Song and W. Zhuang, "QoS Provisioning via Admission Control in Cellular/wireless LAN Interworking," in Proc. of BROADNETS, pp. 585-592, Oct 2005.

[26] X. Gelabert, J. Pérez-Romero, O. Sallent, and R. Agustí, "A Markovian Approach to Radio Access Technology Selection in Heterogeneous Multiaccess/multiservice Wireless Networks," IEEE Transactions on Mobile Computing, Vol. 7, No. 10, pp. 1257-1270, Oct 2008.

[27] A. Roy and A. Karandikar "Optimal Radio Access Technology Selection Policy for LTE-WiFi Network," in proc. of WiOpt, pp. 291 - 298, May 2015.

[28] E. Altman, T. Jiménez, and G. Koole, "On Optimal Call Admission Control in a Resource-Sharing System," IEEE Transactions on Communications, Vol. 49, No. 9, pp. 1659-1668, Sep 2001.

[29] G.A. Brouns and J. Van Der Wal, "Optimal threshold policies in a two-class preemptive priority queue with admission and termination control," Queueing Systems, Vol. 54, No. 1, pp. 21-33, 2006.

[30] A. Turhan, M. Alanyali, and D. Starobinski, "Optimal admission control in two-class preemptive loss systems," Operations Research Letters, Vol. 40, No. 6, pp. 510-515, 2012.

[31] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile Data Offloading: How Much Can WiFi Deliver?," IEEE/ACM Transactions on Networking, Vol. 21, No. 2, pp. 536-550, Apr 2013.

[32] N. Wang and J. Wu, "Opportunistic wifi offloading in a vehicular environment: Waiting or downloading now?," in proc. of Computer Communications, IEEE INFOCOM, pp. 1-9, Apr 2016.

[33] S. Ranjan, N. Akhtar, M. Mehta, and A. Karandikar, "User-based Integrated Offloading Approach for 3GPP LTE-WLAN Network," in proc. of NCC, pp. 1-6, Feb 2014.

[34] B. H. Jung, N. Song, and D.K. Sung, "A Network-assisted User-centric WiFi-Offloading Model for Maximizing Per-user Throughput in a Heterogeneous Network," IEEE Transactions on Vehicular Technology, Vol. 63, Issue 99, pp. 1940 - 1945, Oct 2013.

[35] T. Bonald and J. W. Roberts, "Internet and the Erlang formula," ACM SIGCOMM Computer Communication Review, Vol. 42, No. 1, pp. 23-30, Jan 2012.

[36] A. Roy, P. Chaporkar, and A. Karandikar "Optimal Radio Access Technology Selection Algorithm for LTE-WiFi Network," Department of Electrical Engineering, Indian Institute of Technology Bombay, Tech. Rep., Available online at https://www.ee.iitb.ac.in/course/~arghyadip/LTE-WiFi.pdf, May 2017.

[37] E. Altman, Constrained Markov decision processes, CRC Press, 1999.

[38] D. V. Djonin and V. Krishnamurthy, "Q-Learning Algorithms for Constrained Markov Decision Processes With Randomized Monotone Policies: Application to MIMO Transmission Control," IEEE Transactions on Signal Processing, Vol. 55, No. 5, pp. 2170-2181, May 2007.

[39] 3GPP TR 36.814 V9.0.0, "Further Advancements for E-UTRA Physical Layer Aspects," Available online at http://www.3gpp.org/dynareport/36814.htm, Mar 2010.

[40] 3GPP TR 36.839 V11.1.0, "Mobility Enhancements in Heterogeneous Networks," Available online at http://www.3gpp.org/dynareport/36839.htm, Dec 2012.